Problem: Energy Modeling and Structural Analysis of Gold Clusters

Metal clusters, especially gold (Au) clusters, are tiny groups of atoms with unique properties that depend on their size and arrangement. They are neither single atoms nor bulk materials, and their behavior is influenced by special quantum effects. Even clusters with the same number of atoms can have very different 3D shapes (called configurations or isomers). Each shape has a specific energy, and usually, the one with the lowest energy is the most stable. Figure 1 shows several examples of Au_{20} (20-atom gold) clusters with different shapes.
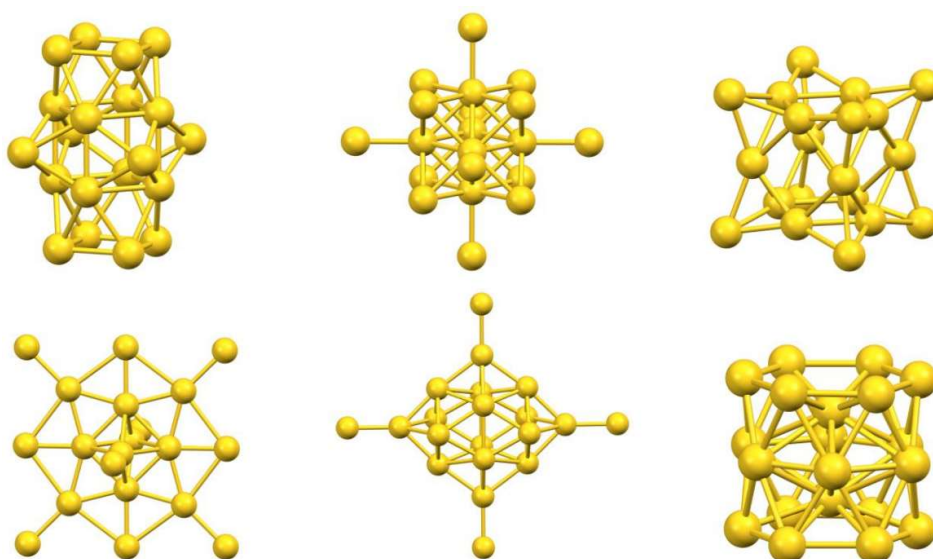

Figure 1. Illustrative representations of various isomeric configurations of the Au_{20} cluster.

One accurate way to calculate these energies is Density Functional Theory (DFT). However, DFT calculations take a lot of computer time, especially for larger clusters. That's why researchers are interested in building faster models—such as mathematical formulas or machine learning models—that can predict the energy of a cluster directly from its atomic coordinates.

To accomplish this task, you will work with a dataset containing 1000 Au_{20} clusters in .xyz format.
- **First line** – Number of atoms (e.g., 20)
- **Second line** – Total energy of the structure (a single number)
- **Remaining lines** – The (x, y, z) coordinates of each atom

You can open .xyz files with a text editor or view them in molecular visualization software like Visual Molecular Dynamics (VMD).

Tasks

Task 1: Predicting Energies of Au_{20} Clusters

You are provided with a dataset containing the three-dimensional coordinates and corresponding total energies of 999 Au_{20} (gold-20) cluster structures.

- Develop a mathematical or machine learning-based model to predict the energy of Au_{20} clusters from their atomic coordinates.
- Use Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 score to evaluate the prediction accuracy of your model.

Task 2: Finding and Describing the Most Stable Structures

Using the same dataset of 999 Au_{20} clusters:

- Analyze the statistical distribution of total energies, including key metrics such as the mean, variance, and skewness.
- Identify and visualize the one of the lowest-energy structures.
- Summarize common geometrical features of low-energy configurations

Task 3: Sensitivity Analysis via Local Structural Perturbation

Using the identified representative lowest-energy Au_{20} cluster structure in Task 2:

- Apply the local perturbations (e.g., random displacements of a few atomic positions) to generate a set of new configurations (slightly "distorted" versions).
- Quantify the relationship between perturbation magnitude and corresponding changes in total energy.
- Use your model in Task 1 to predict the energies of these new versions, and report the changes in predicted energy using MAE and RMSE relative to the original structure.
- Propose a structural stability variable to investigate the robustness of atomic configuration to perturbations.